BioscienceHorizons Volume 10 2017 10.1093/biohorizons/hzx017

Review article

Myth and fact in the origins of cellular life on Earth

Paul Jowett*, Richard Rayne and Salvador Tomas

Department of Biological Sciences, Birkbeck, University of London, Malet Street, London WC1E 7HX, UK

***Corresponding author:** Thessaly Cottage, Foxcombe Lane, Boars Hill, Oxford OX1 5DH, UK. Tel: +44-1865 321107. Email: paul.jowett@rolandberger.com

Supervisor: Salvador Thomas, Department of Biological Sciences, Birkbeck, University of London, Malet Street, London WC1E 7HX, UK

Research into the origins of the first *protocell* is of a multidisciplinary nature. It draws evidence from what we know about the Earth's early atmosphere and environment, and about the most ancient features of the cell's structure and composition. Such data provides the input for the hypothesis generation and experimental reconstruction necessary to mimic steps in the formation of the first protocell. While research into the origins of the first protocell is condemned to focus upon laboratory experiments, it should be guided by a detailed study of real evidence pertaining to the environment on Earth 4 billion years ago. In this review, we take stock of the research that has been performed to date across the main disciplines of earth sciences, biochemistry, and molecular biology. We seek to identify the progress made in laying down a sequence for the events that led up to the first protocell. We also assess the strengths and weaknesses of the experimental designs and suggest some future approaches. While the field has made many important advances, from the original Stanley Miller experiment establishing 'life from chemistry' products such as amino acids, through to Deamer's findings on fatty acid membranes and Szostack's work on lipids, there is still a long and challenging journey ahead to understand how cellular life began. The experiments required to make more rapid progress in the field will likely be more elaborate, costly, and time consuming.

Key words: abiogenesis, prebiotic, self-assembly, protocell, evolution, RNA

Submitted on 14 November 2016; editorial decision on 6 October 2017

Introduction

While at first sight the task of tracing the origins of primitive cellular life on Earth would appear to be the historian's domain, the lack of sources forces us to imagine and then reconstruct the events that could have happened, approaches no historian would employ. The paucity of fossil evidence makes it essential to follow the path of hypothesis generation and experimental reconstruction to shed light on the origins of the first protocell (see Schopf, 1978, 1993, 2006, for attempts to gain fossil evidence, Fig. 1 for the putative results thereof, and see Brasier *et al.*, 2002 for a critique of what has been found). Our theories for how life first evolved on Earth must be built upon laboratory research. This paper will review our current understanding of what constitutes the essentials for cellular life, and track progress made in mimicking the path by which cells came into existence. We will summarize some of the current theories on how cellular life arose, examine the key experiments that have yielded the best results to date, and take stock of the current experimental designs and their limitations.

We will also identify the challenges confronted in improving our theories on the origins of the cell, which include: (1) the possibility that many of the early nucleic and amino acids shaping the first protocell became extinct; (2) the obstacles to reconstructing the environment of 4–3.5 billion years ago and (3) the uncertainty over whether there was one original protocell, the result of a unique sequence of events, or many different protocells, that developed spontaneously on many occasions once the appropriate environmental conditions were appropriate.

Era	Hadean	Archean		Proterozoic				P-M-C
Time	4.5 Ga	4.0 Ga	3.5 Ga	3.0 Ga	2.5 Ga	2.0 Ga	1.5 Ga	1.0 Ga	0.5 Ga
Evidence	Age of moon & meteorites	Oldest reported igneous rocks Oldest sedimentary rocks	Oldest fossil-like objects Oldest carbon with isotope ratio suggesting photo-synthesis	Oldest stromat-olites	First wide-spread diverse stromat-olites Oldest micro-fossils Major banded iron formations	Youngest detrital uraninites	Increase in diversity of micro-fossils Increase in size of spheroidal micro-fossils	Decrease in abundance & diversity of stromat-olites Oldest large algae	Oldest fossils of invert-ebrates
Interpre-tation	Earth's Formation	Ocean & Continent Formation	Inorganic release of traces of oxygen	First anaerobic bacteria	Great oxygenation event Origin of photo-synthetic bacteria Anaerobic prokaryote diversifica-tion	Aerobic photo-synthesis	Oxygen rich atmosphere Aerobic respiration develops Some anaerobic extinctions Eukaryote origins	Hard parts in algae & metazoans Origin of mega-scopic eukaryotes Sexual repro-duction	Large eukaryotic organisms

Figure 1. Putative timing of evolution. In his article in 1978, Schopf outlined the major events based on his and colleagues' assessments of the fossil record and provided tentative estimates of the timing of major events in the evolution of life. P-M-C = Palaeozoic, Mesozoic, Cenozoic. Adapted from Schopf (1978) The evolution of the first cells. *Scientific American* 239(2):110–38

Discussion

The essentials for cellular life

While popular definitions of life, heavily influenced by Darwin, centre on reproduction and the capacity for evolution, biologists in recent years have shifted their focus to the role played by autopoiesis, homoeostasis, cognition, and co-emergence. Autopoiesis is defined as the ability of the cell to replenish and renew itself. Homoeostasis is the ability of the cell to maintain key parameters within the set boundaries required for function. Cognition implies that the cell is able to sense its own needs within its surroundings. Co-emergence refers to the interplay between the cell and its environment. Current definitions of life as expressed by some biologists assume that the cell is more than the sum of its parts: all of the cell's components are dependent upon one another (Luisi, 2014). This observation is particularly pertinent for the chicken-and-egg debate over the sequence of events leading to the self-assembly and self-replication of the constituent components of the protocell: no self-replication of any cell component has ever been achieved in isolation, and the alternative may be that self-replication is impossible on a molecular self-sufficient basis, but rather dependent upon the interplay of disparate molecular assemblies (Segré and Lancet, 2000).

The early hypotheses on the origins of the protocell

The central hypothesis on the origin of the cell is that organic molecules self-assembled to form the first protocell some 4 billion years ago (Fig. 2). Attention has focussed upon four critical processes: the formation of organic molecules such as amino and nucleic acids, the polymerization of these molecules, the formation of membranes, and the development of metabolic networks for power. Competing theories exist for how each of these processes evolved, and in what sequence. While there is a consensus that organic molecules came first, since nucleic and amino acids are the essential building blocks of life, views differ over whether metabolism, polymerized molecules or membranes then followed.

Darwin originally suggested that life began in a 'warm little pond' (Darwin, 1859), but it was Oparin in the early 20th Century who conceived of a model whereby simple molecules, subjected to chemical reactions, became organic, underwent further reactions to assemble macromolecules and polymer complexes, and harnessed metabolic pathways to forge the first cell (Oparin, 1924). The 'life from chemistry' model was a vital first step in the formation of current views on the cell's origins.

Both Oparin and Haldane independently hypothesized that the Earth's initial atmosphere was reductive, rather than oxidative as it is today and thus more likely to provide the metabolic basis for the formation of the organic chemicals essential for life. Oparin defined the term 'primordial soup' to describe the Archean ocean in which solar energy and a reducing atmosphere created organic molecules (Oparin, 1924; Haldane, 1929) (Fig. 3).

The debate over the order of the remaining three events has been lively. Wächtershäuser, a German patent lawyer

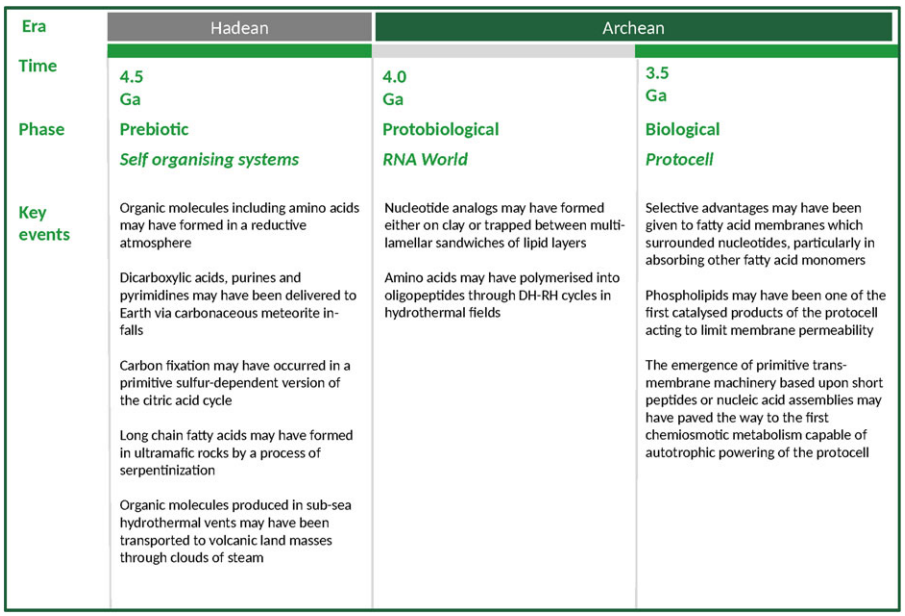

Figure 2. Putative process of protocell development. In his review article of 2005, Pereto set out a scheme for a hypothetical transition from prebiotic chemistry to protocells. Pereto chose not to add a timescale to his scheme. Adapted, with key events defined by author, from Pereto (2005) Controversies on the origin of life. *International Microbiology* 8:23–31

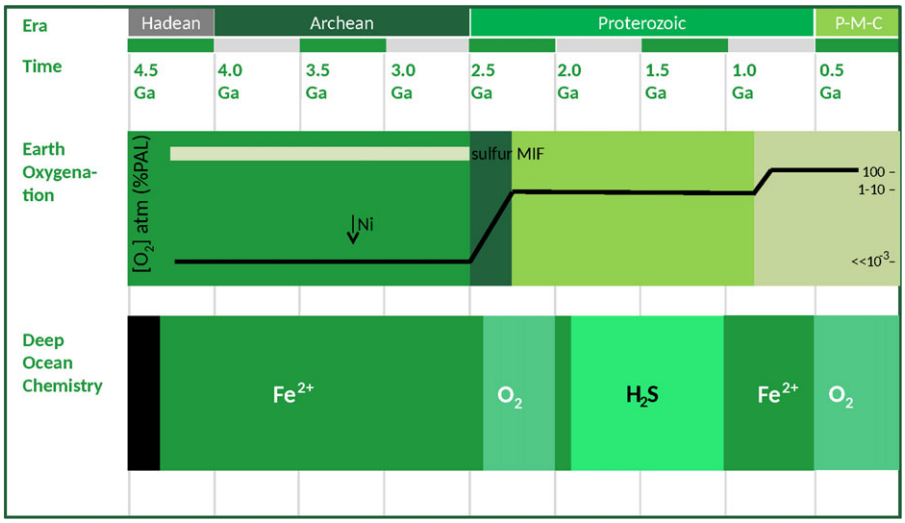

Figure 3. Environmental influences relevant to protocell development. In their review article (2012) Pufahl and Hiatt collated a broad range of sources on environmental change that could impact the biosphere. They stress that the level of oxygenation immediately after the Great Oxidation Event is still largely unknown. Earth oxygenation is divided into three stages in their schema (1,2,3). PAL = present atmospheric levels, MIF = mass-independent fractionation. Adapted from Pufahl and Hiatt (2012) Oxygenation of the Earth's atmosphere-ocean system: a review of physical and classical sedimentologic responses. *Marine and Petroleum Geology* 32:1–20

turned evolutionary biologist, argued that metabolism came first with his 'iron-sulphur world' theory (Wächtershäuser, 1990). Orgel and Gilbert argued equally passionately that the polymerization of RNA took priority, putting the 'RNA World' at the head of the pack (Orgel, 1968; Gilbert 1986). And finally, Szostak, Deamer, and Luisi have put forward arguments for membranes being paramount, in part because they provide the spatial localization necessary to promote non-enzymatic polymerization (Szostak, Bartel, Luisi, 2001; Luisi, 2014; Deamer, 2017).

Early experiments on the formation of organic molecules

The most significant experiment performed to date on the formation of organic molecules remains that of Miller and Urey in 1952 when they sought to test the hypothesis of Oparin and Haldane that the organic building blocks of life emerged from a reducing atmosphere (Miller, 1953). Miller filled a flask with the four gaseous elements of the reductive atmosphere: methane, ammonia, hydrogen and water. He then passed electrical discharges through them to simulate lightning. After 125 h of the experiment, using the equipment available to him at the time, Miller identified several alpha amino acids contained within the residues.

Later analyses of the outputs of Miller's experiment, using more advanced equipment conducted by Bada, revealed that they included a much broader range of amino acids than had originally been detected (Bada, 2009).

Similar progress has been achieved in identifying paths to the formation of long chain fatty acids, such as seen in the work of McCollum, Ritter and Simoneit (1999) that used oxidation of olivine in ultramafic rock by a process of 'serpentinization' to reduce water to molecular hydrogen and subsequently combine it with CO_2 via Fischer-Tropsch synthesis under high temperatures (Segré *et al.* 2001). Potential examples of this process have been found in the Rainbow hydrothermal field on the Mid-Atlantic Ridge (Konn *et al.*, 2015).

By contrast, progress in explaining the formation of nucleic acids is much less well advanced. Indeed Hud and Cafferty (2013) argue that given the failure of all attempts to mimic an abiotic origin for RNA, they feel the evidence suggests that RNA is the product of molecular ancestors, not their forebear. Hud and Cafferty base their views on a series of limiting factors influencing the three distinct molecular units of RNA (the nucleobase, the ribose sugar, and the phosphate group). To these, they add cations and solvent molecules as additional vital components. Each component faces severe abiotic challenges. Researchers have grappled with the problem of how activated pyrimidine ribonucleotides could have synthesized in prebiotically plausible conditions, most notably through the work of Powner, Gerland, and Sutherland (Powner, Gerland, Sutherland, 2009). Powner *et al.* considered the formation of ribonucleotides from their constituent parts too difficult to replicate and advanced an alternative theory, validated by experiments, that ribonucleotides could undergo synthesis in a process that bypassed ribose.

Challenges not yet surmounted, such as the prebiotic route to ribose synthesis and the identity of the cations available in the atmosphere, have led Hud and Cafferty (2013) to postulate that RNA did not self-assemble but is rather the covalent merger of at least two different earlier polymers. Indeed it has been proposed that current RNA chemistry evolved from a complex 'gemisch' of nucleotide analogues (Segré and Lancet, 2000).

Perhaps the best clues will continue to be better data on Earth's early environment. For example, computer simulation experiments by Bernath and others using data from a Canadian satellite mission for sensing of the Earth's atmosphere suggest that the Earth's Hadean atmosphere could have contained up to 40% hydrogen, a figure suggestive of an environment even more conducive of the formation of prebiotic organic molecules than the atmosphere used by Miller in his earlier experiment (Bernath *et al.*, 2005).

Extra-terrestrial influences on supply of organic molecules on Earth

Research into the Murchison meteorite, which fell to Earth in 1969 in Australia, has stimulated growth in the field of astrobiology, a field that studies how extra-terrestrial influences could have played a role in the origins of life on Earth and elsewhere in the universe.

A wide range of analyses have been made of the organic compounds contained in the Murchison meteorite (Yuen *et al.*, 1984; Engel, Macko, Silfer, 1990; Pizzarello, Huang, Fuller, 2004; Huang *et al.*, 2005), but perhaps the most comprehensive and conclusive to date is that of a team led by Martins which used compound-specific carbon isotope data to measure purine and pyrimidine compounds. Martins and others found that dicarboxylic acids were the most abundant class of compounds contained within the Murchison meteorite, and many purines and pyrimidines were also found. The results for uracil and xanthine showed positive $\delta^{13}C$ values, uracil being abundant in modern biochemistry, while xanthine is more limited, functioning as an intermediate in the biosynthesis of guanosine and uric acid (Martins *et al.*, 2008).

Proponents of extra-terrestrial influences on prebiotic development point to the carbonaceous meteorite in-falls during the early Hadean phase of Earth's history, and suggest that nucleobases delivered together with sugar-related species and amino acids may have played a critical role in the founding of the RNA world (Engel, Macko, Silfer, 1990; Martins *et al.*, 2008). The chemical building blocks for the 'life from chemistry' thesis could have been formed on Earth, on extra-terrestrial planets and meteorites, or from a combination of the two. There are few current clues as to which source came 'first' or played the dominant role. What is remarkable, however, is that support for the 'life from chemistry' thesis is not restricted to Earth, giving first insights into the potential for life based on common building blocks beyond our planet.

The 'Metabolism First' approach and the 'Iron-Sulphur World'

The 'metabolism first' or 'iron-sulphur world' model was initially theorized without recourse to experimental testing. Wächtershäuser conceived the hypothesis that the early steps towards polymerization and cell formation began in a hot (100 degrees centigrade) and high pressure (several kilometres under the sea) iron-sulphur hydrothermal environment, a view very different from the warm pond model suggested by Darwin. Wächtershäuser proposed that catalytic centres

supported pathways to autocatalytic carbon fixation in a primitive sulphur-dependent version of the citric acid cycle (Wächtershäuser, 1990). Wächtershäuser's theory gained support after the discovery of sub-ocean hydrothermal vents, and Koonin and Martin proposed a variant on Wächtershäuser's, by which energy is supplied by proton gradients within small inorganic cavities deep on the ocean floor (Koonin and Martin, 2005).

MacLeod *et al.* (1994) have proposed that iron sulphide bubbles containing alkaline emerging from hydrothermal vents could have inflated hydrostatically, as witnessed at the Lost City hydrothermal vents near the Mid-Atlantic ridge. Schock (1990); Schock, McCollum, Schulte (1998) have conducted experiments on the chemical equilibrium of H_2/CO_2 which provides conditions for the synthesis of reduced carbon compounds such as methanogens and acetogens. These are critical given their role in synthesizing adenosine triphosphate with the help of the Wood-Ljungdahl acetyl-CoA pathway of CO_2 fixation (Fuchs, 1986).

Russell and Hall (1997), Russell and Martin (2004) argue that microporous cavities within the hydrothermal vents could provide the spatial localization necessary to permit product accumulation and concentration, both vital for self-replication systems to arise. They also suggest that the bases which make up RNA could have functioned as catalysts with co-factors such as FeS for methyl synthesis, a role subsequently played by proteins. Research has followed focusing on experiments with a wide variety of inorganic catalysering co-factors that could accept electrons in the reduction of CO_2 into carbon compounds.

Despite the support gathered by the hydrothermal vents theory, the 'concentration problem' (never fully resolved by the microporous inorganic cavities model) coupled with the challenges of molecule formation and polymerization occurring in salt water, led to an alternative school of thought gaining traction which locates the site of activity in clay-lined fresh-water pools of volcanic land masses (frequently referred to as hydrothermal fields) (Deamer and Georgiou, 2015). Whereas the metabolism theories of the proponents of sub-ocean hydrothermal vents focus upon a single interface between the solid mineral surface of inorganic cavities and the aqueous ocean, the hydrothermal field supporters see advantages in more complex interfaces between atmosphere/water, atmosphere/mineral and mineral/water, which can leverage dehydration/rehydration (DH–RH) cycles to power molecule formation and polymerization (Rajamani *et al.*, 2008; Deamer, 2017).

Experimental designs to test the hydrothermal vent theories have been hampered by the challenges of simulating high pressures and temperatures in the laboratory. By contrast, fewer challenges have stood in the way of laboratory simulation of hydrothermal fields, and research to establish the impacts of evaporation and precipitation upon the processes of molecular self-assembly has been more fruitful. For example, the role of wet–dry cycles at high temperatures in promoting polymerization of amino acids into oligopeptides was demonstrated by Forsyth *et al.* (2015). The three major advantages which have emerged from research to date in support of the hydrothermal fields theory are: (1) the wet–dry cycles allow for increasingly complex interactions; (2) the confines of the rock pools permit product accumulation and (3) light is a potential source of energy, and makes possible a role for light-absorbing-pigments to power the reactions.

Such counter-intuitive ideas of life beginning in very hot conditions, rather than in the typical temperatures of life today, gained further support through Woese's research in the 1970s, which identified *archaea* as a distinct domain, now found in extreme environments which may reflect its origins (Woese, 1967).

The 'Polymerization of RNA First' approach and the 'RNA World'

The 'polymerization of RNA first' thesis has also attracted many proponents. Oparin in his early model assumed that the formation of macromolecules and polymer complexes would follow directly from the spawning of organic molecules on early Earth. Orgel and others hypothesized that RNA preceded both DNA and proteins before the formation of a membrane in the development of the first cell, although such compartments need not have been mineral in nature (Orgel, 2004; Ralser, 2014). Deamer and Szostak have shown that proton gradients arise readily in lipid vesicles (Deamer, 1992; Chen and Szostak, 2004). The work of Cech and Altman identifying the enzymatic property of RNA in the ribozyme triggered Gilbert to see a role for RNA in self-replication (Gilbert, 1986; Westheimer, 1986; Cech, 2002). Debates have since centred on: (a) whether RNA was preceded by some other nucleic acid easier to polymerize; (b) whether RNA or its predecessor could have polymerized in an unenclosed environment and (c) how RNA could spontaneously arise at all (Robertson and Joyce, 2012).

While the hypothesis that cellular life began from an 'RNA World' is widely held for key steps in the origins of the cell sequence, all experiments have so far failed to mimic the polymerization of nucleotides to form RNA. Groups like those led by Orgel, Eschenmoser, Szostak and others have resorted to experimenting with alternative nucleotides to examine whether more reactive molecules could achieve a better performance without the aid of enzymes. Indeed, as previously mentioned, a school of thought holds that RNA had predecessors, and is itself a product of evolution.

The pre-requisite for polymerized oligonucleotides is that a prebiotic stock of random sequence RNAs was available. Should one of these be a triple stem-loop structure of some 40–60 nucleotides, it might function as a replicase ribozyme, and if this replicase ribozyme is capable of replicating itself with about 90% fidelity, a plausible pathway for RNA evolution might be found (Robertson and Joyce, 2012). Such evidence, however, is sparse, and it is currently a very optimistic

view that RNA replicase could emerge from a random stock of oligonucleotides. The question also remains of whether a replicase could act on sequences similar to itself while ignoring unrelated sequences. Researchers have attempted to address these challenges by experimenting with the segregation of specific molecules or clonal lines by aligning them along the grains of mineral clay surfaces or on the surface of fatty acid or lipid membranes. For example, Ferris has pointed to a potential role for montmorillonite clay in catalysing the formation of RNA, having demonstrated that artificially activated nucleotides do polymerize on certain clays, a process considered likely to have taken place before lipid membranes, but potentially within inorganic 'cells' (Ferris *et al.*, 1996; Ferris, 2006). Brasier has suggested pumice as being a rival substrate for life (Brasier *et al.*, 2011).

Bartel and Szostak (1993) succeeded in evolving an RNA ligase ribozyme from a large population of random sequence RNAs. But pessimists point to the fact that there is no known polymerase ribozyme that can sustain its own replication, and argue that therefore no such ribozyme is possible (Robertson and Joyce, 2012). Nevertheless, researchers have demonstrated a robust reaction system for RNA-catalysed RNA replication using a pair of cross-replicating ligase ribozymes (Lincoln and Joyce, 2009). Rajamani and others showed that unactivated mononucleotides polymerize when dried in the presence of a lipid although the conditions cause some loss of the bases (Rajamani *et al.*, 2008; Mungi and Rajamani, 2015). The solution may once again be that no single oligonucleotide prevailed in RNA evolution, but rather that RNA evolution is the result of a complex interplay of many interdependent oligonucleotides, none of which could function alone (Kaufman, 1993; Higgs and Lehman, 2015). Even assuming such polymerase/ligase activity could have evolved independently from cells, there remains the question of how sufficient feedstock would have been available from the starting materials abundant on Earth 4 billion years ago.

The lack of progress in polymerizing RNA in isolation may be evidence that RNA was not the first nucleotide in the formation of the cell, and indeed the notion of the 'RNA World' has been described by Bernhardt as 'the worst theory of the early evolution of life (except for all the others)' (Bernhardt, 2012). Many candidates have been proposed as alternatives to RNA for the feedstock of oligonucleotide polymerization experiments. Eschenmoser (1999) and others have performed studies on an extensive range of nucleic acid analogues. Their findings suggested that pyranosyl RNA made a good candidate, given its reaction properties, but this candidacy was short-lived given its inability to base pair with RNA (Robertson and Joyce, 2012). Threose nucleotides (TNAs) were found to pair with RNA, and the ability for an analogue to base-pair with RNA has since become an important selection criteria. Orgel (1989) and others have worked on peptide nucleic acid (PNA) which has a backbone held together by amide bonds, and can transfer information from PNA to RNA or vice versa in template-directed reactions. There is credible evidence therefore that a transition was possible from a PNA to an RNA world. It would require more information about the compounds available on primitive Earth to determine whether p-RNA, TNA, GNA, PNA, ANA, tPNA or some other analogue is the best candidate to precede RNA (Robertson and Joyce, 2012).

Despite the many seemingly insuperable obstacles to validating the theory, the 'RNA World' school has continued to attract significant research interest over the years, supported by the likes of Gilbert, Crick, Woese and others.

The 'Membranes First' approach and the 'Lipid World'

Many theories concerning the polymerization of RNA consider spatial localization to play a key role. The hypothesis that spatial localization was provided initially by inorganic cavities in hydrothermal vents commanded much early attention, but in recent years proponents of the 'membrane first' school have expanded this hypothesis to include a role for lipid vesicles, shifting the focus to hydrothermal fields as the more likely location of molecular formation and polymerization.

The key factors underpinning the view that the membrane came first in the evolution of the protocell include: (a) the requirement for spatial localization to achieve product accumulation and concentration; (b) the benefits for prebiotic non-enzymatic polymerization of monomers trapped within multi-lamellar sandwiches of lipid layers; (c) the potential for rate enhancing chemical reactions enhanced by the presence of certain lipid vesicles referred to by Segré *et al.* (2001) as lipozymes (Fendler and Fendler, 1975; Cuccovia, Quina, Chaimovich, 1982; Luisi, Walde, Oberholzer, 1999); (d) the advantages for condensation cycles, involving the loss of water molecules, of separating molecules by means of a membrane from an aqueous environment; (e) the protection also afforded to growing polymers from hydrolysis; (f) the barriers to free diffusion of monovalent cations afforded by membranes, permitting their capture and use in the catalysis of polymerization; and (g) the selection advantages that would accrue to protocells that develop machinery for influencing transmembrane transport of molecules (Deamer, 2017).

While hydrothermal fields dominate current research into the role of membranes, not all proponents of the primacy of lipid membranes have dismissed a role for hydrothermal vents. Grochmal *et al.* have postulated that porous fatty acid walls could develop within vent pours which would permit absorption of nucleotides and/or peptides and promote non-enzymatic oligomerisation reactions (Grochmal *et al.*, 2015). Lane argues that lipids might have formed within the vent pours lining them with leaky lipid membranes (Lane, 2012).

Without a membrane, it is difficult to conceive of an autonomous cell, and given that amphiphilic lipids automatically self-assemble into vesicles when placed in water, it is easy to imagine them providing a viable membrane to the first

protocells. The first protocell walls were likely porous, composed of simple single-chain fatty acids capable of being both synthesized in the environment and transported to Earth on meteorites. Budin and Szostak have investigated the evolutionary importance of lipid membranes in establishing selective advantages. Their theory is that the first RNA machinery may have produced phospholipids, which when added to the fatty acids in the membrane reduced membrane permeability. For this reason they believe the second stage of RNA evolution included the development of transmembrane transport machinery (Budin and Szostak, 2011).

The earliest protocells will have had very little functionality of any kind, and instead responded largely passively to environmental influences for their operation. Fatty acid membranes are highly permeable, which is consistent with a heterotrophic model by which chemical building blocks synthesized outside the protocell are diffused passively across the cell membrane to contribute to functions that confer selective advantages. Any increase in the sophistication of the membrane wall, permitting the cell to retain internally synthesized metabolites, would be dependent upon an increase in the sophistication of the machinery established within the protocell. Szostak (2011) and his collaborators have shown that selective advantages are given to fatty acid protocell membranes surrounding RNA as they absorb fatty acid monomers from other empty membranes or micelles. In their most recent work they found that small increases in phospholipid content lead to a cascade of additional selective pressures for transmembrane transport machinery that might be produced by primitive acyltransferase ribozymes. The early transporters are thought to have been formed by short peptides or nucleic acid assemblies (Szostak, 2011).

The development of lipid membranes is highly synergistic with the development of early catalysts such as phospho- and acyltransferases, which once established could quickly be adapted for use in metabolic tasks such as sugar catabolism and peptide synthesis (Budin and Szostak, 2011). Furthermore, the increasing phospholipid content of the membrane would enable the emerging protocell to utilize transmembrane ion gradients, the basis for the chemiosmosis vital for metabolism in all extant cells (Szostak, Bartel, Luisi, 2001).

Template-directed non-enzymatic replication and reproduction

The model of the first protocell, in which little machinery is present and the protocell responds passively to environmental influences for its function, sits poorly with a major role being given to template-directed replication and reproduction (Deamer, 2008), particularly in comparison with the much more simple role of compositional information contained in assemblies of amphiphilic molecules (Segré *et al.*, 2001). For evolutionarily selective advantages to be established it is essential that protocells be able to divide and pass on their information and structures to their daughters. While DNA-coded protein synthesis is now the basis of replication and reproduction, considerable evidence supports the hypothesis that early genetics was largely the province of RNA. Szostak (2012) has divided the 'RNA World' into two phases, the first centring on the prebiotic synthesis of ribonucleotides, and the second encompassing template-directed replication of RNA molecules within replicating protocells. The Sutherland lab helped define ways in which the first phase could have been achieved (Powner, Gerland, Sutherland, 2009), but progress on the second phase is much slower and more challenging. The principle hypothesis currently being tested is that the second phase involved non-enzymatic RNA replication with the implication that the first heritable RNA might have been a metabolic ribozyme that conferred enhanced protocell reproduction or survival (Szostak, 2011). This is a minimalist approach to reproduction, as it assumes no machinery is put in place to improve replication, and that the fidelity of replication may be very poor.

Szostak (2012) has defined an eight-fold path to non-enzymatic RNA replication in his blueprint for the design of experimental reconstructions which could prove the efficacy of chemically driven RNA replication. But despite the neat division of the challenges into categories, he himself highlights the interdependence and interaction between the potential solutions to each challenge, making RNA replication a truly 'systems chemistry' problem, in which every aspect of a complex system must be considered as part and parcel of the whole.

The principle obstacle non-enzymatic template-directed replication needs to overcome is the outperformance of strand re-annealing relative to non-enzymatic template copying (Jia *et al.*, 2016). The RNA duplex produced by non-enzymatic copying must denature to supply templates for subsequent replication, and research has focussed on mechanisms to identify: (a) properties that would facilitate strand separation and (b) properties which permit non-enzymatic RNA replication to out-compete re-annealing of the duplex RNA strands. A further challenge is that functional RNA sequences need to be able to self-replicate before they degrade. Many in the Szostak lab and elsewhere have attempted to overcome these challenges, currently without material success. Jia *et al.* (2016) did believe that the solution to the re-annealing problem might be found in arginine-rich oligopeptides that could function as primers both promoting copying and inhibiting re-annealing, but this thesis was quickly withdrawn (Szostak, 2017). More progress has been made in achieving non-enzymatic copying of RNA templates containing all four nucleotides (A,C,G,T) through catalysis of activated oligonucleotides (Prywes *et al.*, 2016; Wachowius, Attwater, Holliger, 2017). The challenges faced for non-enzymatic replication of RNA through requirements for high concentrations of Mg^{2+} (or other divalent) ions and their destructive power over fatty acid membranes, were resolved when Adamala and Szostak (2013) demonstrated that the presence of citrate could inhibit this destructive influence. A possible solution to the problem of the high melting point of RNA has been suggested which would imply that the RNA backbone may have contained both 2' and 3' linkages, which requires less fidelity in reproduction (Bernhardt, 2012). It has also been suggested that the environment played a major role in supporting the

replication process. Varying extremes of temperature would provide an ideal environment for different steps, with cold environments hosting template copying and hot ones strand separation and the influx of nutrients and nucleotides (Budin and Szostak, 2010).

While we are still a long way from achieving a successful experimental reconstruction of early RNA replication, promising work has been done on both strand separation, and the fidelity of the replication process.

The extinction of proteins and implications for their role in the protocell

One of the biggest challenges faced in considering the role played by proteins in the emergence of the protocell is the high probability that many of them may have been lost due to selective pressure from environmental changes. The single biggest change likely to have caused protein extinction is the Great Oxidation Event, when Earth moved from a reductive to an oxidative atmosphere around 2.5 billion years ago (Holland, 2002; Sessions *et al.*, 2009; Shields-Zhou, 2011; Pufahl and Hiatt, 2012).

To date, the theory of parsimonious conservation, by which evolution occurs with the lowest possible number of changes, has had a big influence on consideration of the role played by proteins in the protocell, partly because of its convenience as a hypothesis. The overwhelming challenge posed by rejecting this theory, and accepting that most if not all primordial proteins are now extinct, is that the ratio of potential proteins to known proteins is so large. Even after taking into consideration the arguments proposed by Chiarabelli and De Lucrezia on the one hand, who consider the frequency of stable folds being likely to reduce the potential number, and Marsh on the other, who argues evolution could not have taken place on the basis of an unbiased selection, the number of proteins which could have evolved and then been lost is enormous (Chiarabelli and De Lucrezia, 2007; Marsh, 2013). Rediscovering extinct proteins, and the enzymatic functions they held, is a formidable challenge. Even if we discover potential protein candidates for the cell's early development, it will be impossible to prove that they either existed or played a role.

Despite the very great challenges of identifying the proteins involved in protocell development, our ability to practice experimental reconstruction in the development of proteins has advanced by leaps and bounds. Miller and Gulbis have reviewed the growing tool-kit now available to researchers, including the ability to clone, modify, express, reconstitute and recombine an extensive range of proteins (Miller and Gulbis, 2015). Researchers are also developing technologies that could assist with building protocells *de novo*. Complex protocell designs may soon be able to incorporate machinery for product synthesis, energy generation, protein transcription, translation machinery and the capacity for replication. While simpler versions may dispense with replication capacity, they would likely retain template-guided processes that allow for the renewal of individual components. Miller and Gulbis consider the lack of reproductive capacity renders a simplified protocell unworthy of the label 'life' form. Yet at the dawn of the cell's origins replication may have been achieved through the interplay between protocells and their environment.

The role of the environment in early protocell formation

Recognition that amino acid, nucleic acid and protein extinctions may pose one of our biggest challenges to protocell origins research brings to the fore the most crucial areas for future experiments. The best clues we have currently as to which organic molecules and proteins played an early role in the cell's origins are to be found in the environment in which they evolved. Advances in our knowledge will depend on top down approaches, such as molecular phylogeny, and bottom up approaches, such as palaeobiochemistry (Fig. 4). Oparin and Miller demonstrated the power of combining theories about the conditions on Earth 4 billion years ago with experiments that mimic those conditions. In the absence of cell-specific fossils, we must rely upon clues concerning their environment. Such evidence may yet be found under the beds of deep oceans (Russell and Martin, 2004; Keller, Turchyn, Ralser, 2014; Ralser, 2014), or in rock formations under the polar caps. Such sources may also be found in today's deep biosphere, in conditions analogous to those found on the planet's early days. Or the evidence may lie in meteorites, and samples from other planets. The more light we can cast on the conditions existing on Earth 4 billion years ago, the more hope we have of understanding the early steps in life (Fig. 3). By narrowing down the range of conditions, we may be able to find credible candidates for the first nucleotide and peptide combinations. These candidates should also, of course, undergo back-solving analyses from current cells, and this is an area where sequencing may be of great assistance.

Current assumptions on the Earth's early environment are focusing our attention on *halophilic archaea*, which inhabit extreme environments exposed to heat, salt, and highly acidic or alkaline environments (Woese, Kandler, Wheelis, 1990; Caetano-Anollés, Kim, Mittenthal, 2007; Ellis, Bizzoco, Kelley, 2008). Such *archaea* occupy an environment similar to the hydrothermal vents on the ocean floor. Thermoplasma and Ferroplasma are *archaea* lacking cell walls, living in hot acidic soils. They are considered to have parallels with the protocell functioning within inorganic walls. The fact that these organisms can exist in anaerobic conditions, and power themselves with chemical energy from hydrogen protons, sulphur and iron is considered particularly relevant.

Some of the most promising research conducted recently has examined ways in which protocells might have travelled from one environment, such as hydrothermal vents under the sea, to another, such as trapped clay-lined pools on land. Ellis, Bizzoco, and Kelley experimented with new techniques to overcome the challenges facing collection of sufficient

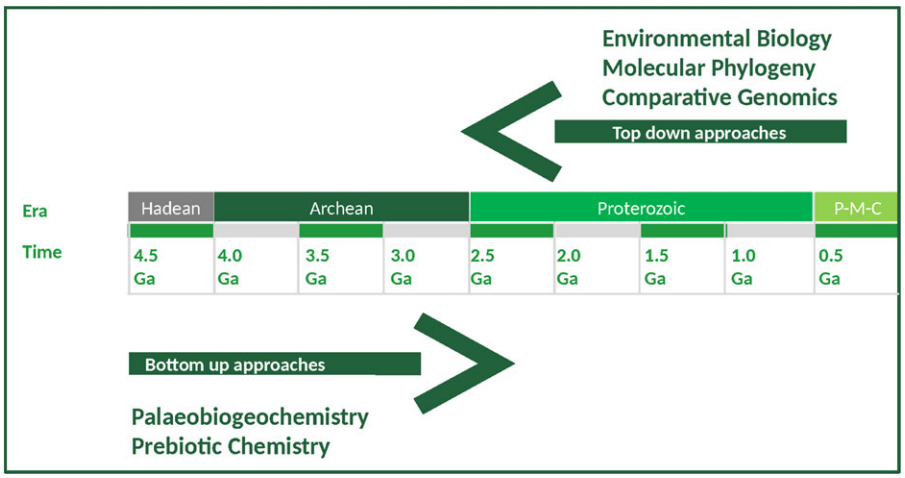

Figure 4. Multi-disciplinary approaches to origins of cellular life research. In their perspective article (2008) Patrick Forterre and Simonetta Gribaldo produced a schematic highlighting the bottom-up and top-down approaches tackling origins of life research and detailing the contribution of different disciplines to the research. Adapted from Forterre and Gribaldo (2008) The origin of modern terrestrial life. *Human Frontier Science Program Journal* 1(3):156–68.

quantities of steam water from hydrothermal vents to examine *archaea* migration. Their success using a novel portable steam collector enabled them to deduce that considerable quantities of microbes from such events could disperse over great distances along with the steam water. This work has been very influential in underpinning the hypothesis that protocells supported by hydrogen sulphide contained in deep faults and fractures could populate broader regions of the planet through air-based dispersion (Ellis, Bizzoco, Kelley, 2008). It forms a vital connection between the contrasting environments of hydrothermal vents and hydrothermal fields.

The metabolic variety of *archaea* gives scope for multiple origins of the cell. The fact that cells may have emerged only 700 million years after the planet was formed is held by many to be proof that they evolved 'as soon as they could', and suggests: (a) an inevitability of life emerging in that environment and (b) life emerging many times rather than once at a unique confluence of events. These insights suggest that in looking at the conditions of life on Earth 4 billion years ago our search should not be restricted to a single environment or event. There is sufficient probability that protocells and their components emerged in multiple environments, i.e. both hydrothermal vents and hydrothermal fields, and that these early cells exchanged genetic material by means of lateral transfer both within and between species. Such views challenge a long-standing Tree of Life dogma, by which all life stems from a single root. In its place a 'web of life' theory has emerged as an alternative (McInerney, Cotton, Pisani, 2008; O'Malley, Martin, Dupre, 2010). We must face the prospect that the objective of our pursuit is not a single original cell from one environment, but multiple original cells from many locations.

The role played by lateral gene transfer in the origin of protocells is also supported by what we now know about viruses. Given that viruses depend upon cells for their reproduction, they cannot predate cellular life, but current consensus views are that they emerged simultaneously with cells, and played an important role in their genetic exchange (Filee, Forterre, Laurent, 2003; Claverie, 2006; Koonin, Senkevich, Dolja, 2006).

Future directions in research

Future research will build on what has already been tried and tested. The hypotheses which have received most attention will continue to be the focus of activity including the theories that: (a) organic molecules including amino acids may have formed in a reductive atmosphere; (b) dicarboxylic acids, purines and pyrimidines may have been delivered to Earth via carbonaceous meteorite in-falls; (c) carbon fixation may have occurred in a primitive sulphur-dependent version of the citric acid cycle; (d) long chain fatty acids may have formed in ultramafic rocks by a process of serpentinization; (e) organic molecules produced in sub-sea hydrothermal vents may have been transported to volcanic land masses through clouds of steam; (f) nucleotide analogues may have formed either on clay or trapped between multi-lamellar sandwiches of lipid layers; (g) amino acids may have polymerized into oligopeptides through DH–RH cycles in hydrothermal fields; (h) selective advantages may have been given to fatty acid membranes which surrounded nucleotides, particularly in absorbing other fatty acid monomers; (i) phospholipids may have been one of the first catalysed products of the protocell acting to limit membrane permeability; and (j) the emergence of primitive transmembrane machinery based upon short peptides or nucleic acid assemblies may have paved the way to the first chemiosmotic metabolism capable of autotrophic powering of the protocell.

Future experiments designed to test these and other hypotheses will likely require more resources devoted to recreating the

early environments in which protocells evolved. To date, the environments used for prebiotic reconstruction experiments have been constrained to laboratory conditions recognized as unconducive for the cultivation of most bacteria. Future experiments will need not only to mimic the environmental conditions in hydrothermal vents and hydrothermal fields of 4 billion years ago, but also to link them, in order that the products of one may freely migrate to mix with the products of the other. The extent to which different organisms are interdependent is frequently underestimated, and in attempting to recreate conditions on primitive Earth, it will be beneficial to facilitate some degree of interdependence within the modelled prebiotic environment.

Ideally the experimental design should be so constructed to ensure that the DH–RH cycle and other processes can be accelerated to achieve results in years rather than millennia. To maximize the chances of achieving a successful result, the experimental design should also be constructed in a modular fashion to permit many different variants to be processed simultaneously as a form of high throughput testing. The monitoring mechanism will also ideally be real-time in order to ensure that positive results in any experiment can be identified immediately, and learnings quickly applied to new experimental designs.

Experiments in the future will likely become more multidisciplinary and more costly, potentially requiring infrastructure on the scale of the large Hadron collider at CERN. One experimental design might be to reconstruct a slice of the Archaean atmosphere, hermetically sealed from our own, including both landmass and a portion of reconstructed ocean. The time-frame of such an experiment would necessarily be long, involving many collaborative groups dedicated to different parts of the evolutionary puzzle.

Despite all the obstacles placed in our way, including the possibility of mass extinctions of nucleotides, amino acids, and proteins, we must not lose sight of the forensic nature of our investigation into the past: any clues might help. Rather like the work that took place in producing the periodic table, our ability to advance knowledge on protocell origins may depend upon identifying more precisely the gaps that exist in our knowledge.

Conclusion

Much work is left to be done to: understand how cellular life began; achieve an approximation of the protocell's development; and establish a theory which is both consistent with the conditions of Earth's early environment and with the evolutionary outcome. The end result will be a hypothesis that will very likely be popularized into a new evidence-based myth on how life began on Earth. The extent to which this hypothesis becomes dogma may depend upon what it teaches us about cellular life today.

Recent advances in mapping the evolution of the first protocell have seen us gain an understanding of how organic molecules formed prior to cellular life. In the next decade, advances may be made in explaining more definitively how these early organic molecules polymerized.

The best support for our account of how cellular life began may well rest upon our ability to create some level of cellular life *de novo*. From the contemporary standpoint, achievement of this goal still seems distant, but underpinned by the recent advances in protein engineering it is possible that more ambitious, expensive, and interdisciplinary experiments on the one hand, and massive whole genome mapping of all extant species on the other, may bring this goal closer.

A review of the authors sourced in this paper reveals that 44% are biologists, 40% are geologists, and 16% are chemists. Narrowing this analysis down to more specific disciplines brings earth sciences to pole position (17%), followed by biochemistry (17%), chemistry (15%), and molecular biology (13%). A holistic approach to the origins of the cell is by definition a multidisciplinary activity and will continue to be so, as the design of experiments becomes larger and more complex. Only a multidisciplinary approach will capture fully the interplay between environment and biosphere, the fossil record and experimental reconstruction, the terrestrial and extra-terrestrial sources of information. Only by treating the origins of the cell enquiry as a 'systems' exercise can progress be made on all fronts, and at a speed that will yield results in our lifetime.

Author biography

Paul Jowett completed a BSc in Biomedicine at Birkbeck, University of London, before moving on to the MSc programme in Genomic Medicine at Barts and the London School of Medicine & Dentistry. He originally studied for a BA in History and Politics at Queen Mary, University of London, before researching for a D. Phil in Politics at Nuffield College, University of Oxford. He is keen to discover ways in which genomics and proteomics can help shed light on the earliest origins of life on Earth. Paul manages the London office of Roland Berger, a strategy consultancy, and is the author (with Francoise Jowett) of *Private Equity*: *The German Experience*, along with four other business and politics books published by Palgrave Macmillan.

Acknowledgements

I would like to thank Dr Salvador Tomas of Birkbeck, University of London for his support and feedback while developing and writing this paper. I would also like to thank Professor Jack Szostak of Harvard Medical School, Professor Gerald Marsh of the Argonne National Laboratory at the University of Chicago, and Geraldine Jowett of King's College London for feedback on earlier drafts of this paper.

Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

References

Adamala, K. and Szostak, J. W. (2013) Nonenzymatic template-directed RNA synthesis inside model protocells, *Science (New York, N.Y.)*, 342 (6162), 1098–1180.

Bada, J., (2009) The results of the re-visiting of the Miller-Urey experiment were broadcast by the BBC in 'The spark of life', a television documentary. *BBC 4*, 26 August 2009.

Bartel, D. P. and Szostak, J. W. (1993) Isolation of new ribozymes from a large pool of random sequences, *Science (New York, N.Y.)*, 261, 1411–1418.

Bernath, P. F., McElroy, C. T., Abrams, M. C. et al. (2005) Atmospheric chemistry experiment (ACE): mission overview, *Geophysical Research Letters*, 32 (L15S01), 1–5.

Bernhardt, H. S. (2012) The RNA world hypothesis, the worst theory of the early evolution of life, except for all the rest, *Biology Direct*, 7 (23), 1–10.

Brasier, M. D., Green, O. R., Jephcoat, A. P. et al. (2002) Questioning the evidence for Earth's oldest fossils, *Nature*, 416 (6876), 76–81.

Brasier, M. D., Matthewman, R., McMahon, S. et al. (2011) Pumice as a remarkable substrate for the origin of life, *Astrobiology*, 11, 725–735.

Budin, I. and Szostak, J. W. (2010) Expanding roles for diverse physical phenomena during the origin of life, *Annual Review of Biophysics*, 39, 245–246.

Budin, I. and Szostak, J. S. (2011) Physical effects underlying the transition from primitive to modern cell membranes, *Proceedings of the National Academy of Sciences*, 108 (13), 5249–5254.

Caetano-Anollés, G., Kim, H. S. and Mittenthal, J. E. (2007) The origin of modern metabolic networks inferred from phylogenomic analysis of protein architecture, *Proceedings of the National Academy of Science*, 104 (22), 9358–9363.

Chen, I. A. and Szostak, J. W. (2004) Membrane growth can generate a transmembrane pH gradient in fatty acid vesicles, *PNAS*, 101 (21), 7965–7970.

Chiarabelli, C. and De Lucrezia, D. (2007) Question 3: the worlds of the prebiotic and never born proteins, *Origins of Life and Evolution of Biospheres*, 37, 357–361.

Claverie, J. M., (2006) Virus evolution, from neglect to centre stage, amidst some confusion. *Structural & Genomic Information Laboratory* CNRS-UPR2589 IBSM: 1–10.

Cuccovia, L. M., Quina, F. H. and Chaimovich, H. (1982) A remarkable enhancement of the rate of ester thiolysis by synthetic and amphiphile vesicles, *Tetrahedron*, 38 (7), 917–920.

Darwin, C. (1859) *The origin of species by means of natural selection*, John Murray, London.

Deamer, D. W. (1992) Polycyclic aromatic hydrocarbons: primitive pigment systems in the prebiotic environment, *Advances in Space Research: The Official Journal of the Committee on Space Research (COSPAR)*, 12 (4), 183–189.

Deamer, D. W. (2008) How leaky were primitive cells? *Nature*, 454, 37–38.

Deamer, D. W. and Georgiou, C. D. (2015) Hydrothermal conditions and the origins of cellular life, *Astrobiology*, 15 (12), 1091–1095.

Deamer, D. W. (2017) The role of lipid membranes in life's origin, *Life (Chicago, IL: 1978)*, 7 (5), 1–17.

Ellis, D. G., Bizzoco, R. W. and Kelley, S. T. (2008) Halophilic archaea determined from geothermal steam vent aerosols, *Environmental Microbiology*, 10 (6), 1582–1590.

Engel, M. H., Macko, S. A. and Silfer, J. A. (1990) Carbon isotope composition of individual amino acids in the Murchison meteorite, *Nature*, 348, 47–49.

Eschenmoser, A. (1999) Chemical etiology of nucleic acid structure, *Science (New York, N.Y.)*, 284, 2118–2124.

Fendler, J. D. and Fendler, E. J. (1975) *Catalysis in Micellar and Macromolecular Systems*, Academic Press, New York.

Ferris, J. P., Hill, A. R., Liu, R. and Orgel, L. E. (1996) Synthesis of long prebiotic oligomers on mineral surfaces, *Nature*, 381, 59–62.

Ferris, J. P. (2006) Montmorillonite-catalysed formation of RNA oligomers: the possible role of catalysis in the origins of life, *Philosophical Transactions of the Royal Society B: Biological Sciences*, 361 (1474), 1777–1786.

Filee, J., Forterre, P. and Laurent, J. (2003) The role played by viruses in the evolution of their hosts: a view based on informational protein phylogenies, *Research in Microbiology*, 154, 237–243.

Forsyth, J. G., Yu, S. S., Mamajanov, I. et al. (2015) Ester-mediated amide bond formation driven by wet-dry cycles: a possible path to polypeptides on the prebiotic Earth, *Angewandte Chemie International Edition*, 54, 9871–9875.

Forterre, P. and Gribaldo, S. (2008) The origin of modern terrestrial life, *Human Frontier Science Program Journal*, 1 (3), 156–168.

Fuchs, G. (1986) CO_2 fixation in acetogenic bacteria: variations on a theme, *FEMS Microbiology Review*, 39, 181–213.

Gilbert, W. (1986) The RNA World, *Nature*, 319, 618.

Grochmal, A., Prout, L., Makin-Taylor, R., Prohens, R. and Tomas, S. (2015) Modulation of reactivity in the cavity of liposomes promotes the formation of peptide bonds, *Journal of the American Chemical Society*, 137 (38), 12269–12275.

Haldane, J. B. S., (1929) The origin of life. *The Rationalist Annual for the Year 1929*, edited by Charles A Watts: 3–10.

Higgs, P. G. and Lehman, N. (2015) The RNA World: molecular cooperation at the origins of life, *Nature Reviews: Genetics*, 16, 7–17.

Holland, H. D. (2002) Volcanic gases, black smokers, and the Great Oxidation Event, *Geochimica et Cosmochimica Acta*, 66 (21), 3811–3826.

Huang, Y., Wang, Y., Alexandre, M. R., Lee, T., Rose-Petruck, C., Fuller, M. and Pizzarello, S. (2005) Molecular and compound-specific isotopic characterization of monocarboxylic acids in carbonaceous meteorites, *Geochimica et Cosmochimica Acta*, 69, 1073–1084.

Hud, N. V. and Cafferty, B. J. (2013) The origin of RNA and 'My Grandfather's Axe', *Chemistry & Biology*, 20, 466–474.

Jia, T. Z., Fahrenbach, A. C., Kamat, N. P. et al. (2016) Oligoarinine peptides slow strand annealing and assist non-enzymatic RNA replication, *Nature Chemistry*, 8 (10), 915–921.

Kaufman, S. A. (1993) *The Origins of Order – Self Organisation and Selection Evolution*, Oxford University Press, New York, Oxford.

Keller, A. K., Turchyn, A. V. and Ralser, M. (2014) Non-enzymatic glycolysis and pentose phosphate pathway-like reactions in a plausible Archaean ocean, *Molecular Systems Biology*, 10 (725), 1–12.

Koonin, E. V. and Martin, W. (2005) On the origin of genomes and cells within inorganic compartments, *Trends in Genetics*, 21 (12), 647–654.

Koonin, E. V., Senkevich, T. G. and Dolja, V. V. (2006) The ancient virus world and evolution of cells, *Biology Direct*, 19, 1–29.

Konn, C., Charlou, J. L., Holm, N. G. et al.. (2015) The production of methane, hydrogen, and organic compounds in ultramafic-hosted hydrothermal vents of the Mid-Atlantic Ridge, *Astrobiology*, 15 (5), 381–399.

Lane, N. (2012) Life – inevitable or fluke, *New Scientist*, 23, 32–37.

Lincoln, T. A. and Joyce, G. F. (2009) Self-sustained replication of an RNA enzyme, *Science*, 323 (5918), 1229–1232.

Luisi, P. L., Walde, P. and Oberholzer, T. (1999) Lipid vesicles as possible intermediates in the origin of life, *Current Opinions in Colloid & Interface Science*, 4, 33–39.

Luisi, P. L. (2014) *The Emergence of Life: From Chemical Origins to Synthetic Biology*, Cambridge University Press, Cambridge, UK.

MacLeod, G., McKeown, C., Hall, A. J. et al.. (1994) Hydrothermal and oceanic pH conditions of possible relevance to the origin of life, *Origins of Life and Evolution of the Biosphere*, 24 (1), 19–41.

Marsh, G. (2013) The problem of the 'prebiotic and never born proteins', *International Journal of Astrobiology*, 12 (1), 94–98.

Martins, Z., Botta, O., Fogel, M. L. et al. (2008) Extra-terrestrial nucleobases in the Murchison meteorite, *Earth and Planetary Science Letters*, 270, 130–136.

McCollum, C., Ritter, G. and Simoneit, B. R. T. (1999) Lipid synthesis under hydrothermal conditions by Fischer-Tropsch-type reactions, *Origins of Life and Evolution of the Biosphere*, 29 (2), 153–166.

McInerney, J. O., Cotton, J. A. and Pisani, D. (2008) The prokaryotic tree of life: past, present… and future? *Trends in Ecology and Evolution*, 23 (5), 281.

Miller, D. S. and Gulbis, J. M. (2015) Engineering *protocells*: prospects for self-assembly and nanoscale production lines, *Life (Chicago, Ill: 1978)*, 5, 1019–1053.

Miller, S. (1953) A production of amino acids under possible primitive Earth conditions, *Science (New York, N.Y.)*, New Series 117 (3046), 528–529.

Mungi, C. V. and Rajamani, S. (2015) Characteristics of RNA-like oligomers from lipid-assisted non-enzymatic synthesis: implications for origins of informational molecules on early Earth, *Life (Chicago, Ill: 1978)*, 5, 65–84.

O'Malley, M. A., Martin, W. and Dupre, J. (2010) The tree of life: introduction to an evolutionary debate, *Biology and Philosophy*, 25, 441–453.

Oparin, A. I. (1924) The origin of life. Translation by Anne Synge, published on-line via valencia.edu.

Orgel, L. E. (1968) Evolution of the genetic apparatus, *Journal of Molecular Biology*, 38 (3), 381–393.

Orgel, L. E. (1989) The origin of polynucleotide-directed protein synthesis, *Journal of Molecular Evolution*, 29 (6), 465–474.

Orgel, L. E. (2004) Prebiotic chemistry and the origin of the RNA world, *Critical Reviews in Biochemistry and Molecular Biology*, 39, 99–123.

Pereto, J. (2005) Controversies on the origin of life, *International Microbiology*, 8, 23–31.

Pizzarello, S., Huang, Y. and Fuller, P. (2004) The carbon isotopic distribution of Murchison amino acids, *Geochimica et Cosmochimica Acta*, 68, 4963–4969.

Powner, M. W., Gerland, B. and Sutherland, J. D. (2009) Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions, *Nature*, 459, 239–242.

Prywes, N., Blain, J. C., Del Frate, F. and Szostak, J. W. (2016) Non-enzymatic copying of RNA templates containing all four letters is catalysed by activated oligonucleotides, *eLIFE*, 17756, 1–14.

Pufahl, P. K. and Hiatt, E. E. (2012) Oxygenation of the Earth's atmosphere-ocean system: a review of physical and chemical sedimentologic responses, *Marine and Petroleum Geology*, 32, 1–20.

Rajamani, S., Vlassov, A, Benner, S et al. (2008) Lipid-assisted synthesis of RNA-like polymers from mononucleotides, *Origins of Life & Evolution of Biospheres*, 38, 57–74.

Ralser, M. (2014) The RNA world and the origin of metabolic enzymes, *Biochemical Society Transactions*, 42, 985–988.

Robertson, M. P. and Joyce, G. F. (2012) The origins of the RNA world, *Cold Spring Harbour Perspectives on Biology*, 4, 1–23.

Russell, M. J. and Martin, W. (2004) The rocky roots of the acetyl-CoA pathway, *Trends in Biochemical Sciences*, 29 (7), 358–363.

Russell, M. J. and Hall, A. J. (1997) The emergence of life from iron monosulfide bubbles at submarine hydrothermal redox and pH front, *Journal of the Geological Society of London*, 154, 377–402.

Schock, E. L. (1990) Geochemical constraints on the origin of organic compounds in hydrothermal systems, *Origins of Life and Evolution of Biospheres*, 20, 331–367.

Schock, E. L., McCollum, T. and Schulte, M. D. (1998) The emergence of metabolism from within hydrothermal systems, in Wiegel J. and Adams M.W.W. (eds), *Thermophiles: The Keys to Molecular Evolution and the Origins of Life*, Taylor & France, London UK, pp. 59–76.

Schopf, J. W. (1978) The evolution of the first cells, *Scientific American*, 239 (3), 110–138.

Schopf, J. W. (1993) Microfossils of the early Archaean apex chert: new evidence of the antiquity of life, *Science (New York, N.Y.)*, 260, 640–646.

Schopf, J. W. (2006) The first billion years: when did life emerge? *Elements*, 2, 229–233.

Segré, D., Ben-Eli, D., Deamer, D. W. and Lancet, D. (2001) The lipid world, *Origins of Life and Evolution of the Biosphere*, 31, 119–145.

Segré, D. and Lancet, D. (2000) Composing life, *EMBO Reports*, 1 (3), 217–222.

Sessions, A. L., Doughty, D. M., Welander, P. V. et al. (2009) The continuing puzzle of the Great Oxidation Event, *Current Biology*, 19 (14), 567–574.

Shields-Zhou, G. (2011) Biogeochemistry toxic Cambrian oceans, *Nature*, 469 (7328), 42–43.

Szostak, J. W. (2011) An optimal degree of physical and chemical heterogeneity for the origin of life? *Philosophical Transactions of the Royal Society B*, 366, 2894–2901.

Szostak, J. W. (2012) The eightfold path to non-enzymatic RNA replication, *Journal of Systems Chemistry*, 3 (2), 1–14.

Szostak, J. W., (2017) Unpublished correspondence between Jack Szostak and Paul Jowett on 9th September 2017.

Szostak, J. W., Bartel, D. P. and Luisi, P. L. (2001) Synthesizing life, *Nature*, 409, 387–390.

Wachowius, F., Attwater, J. and Holliger, P. (2017) Nucleic acids: function and potential for abiogenesis, *Quarterly Review of Biophysics*, 50 (e4), 1–37.

Wächtershäuser, G. (1990) Evolution of the first metabolic cycles, *Proceedings of the National Academy of Sciences*, 87, 200–204.

Westheimer, F. H. (1986) Polyribonucleic acids as enzymes, *Nature*, 319, 534.

Woese, C. R. (1967) *The Genetic Code: The Molecular Basis for Genetic Expression*, Harper & Row, New York.

Woese, C. R., Kandler, O. and Wheelis, M. (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya, *Proceedings of the National Academy of Sciences*, 87 (12), 4576–4579.

Yuen, G., Blair, N., Des Marais, D. J. and Chang, S. (1984) Carbon isotope composition of low molecular weight hydrocarbons and monocarboxylic acids from Murchison meteorite, *Nature*, 307, 252–254.